

DIGITAL CURATION/DIGITAL ARCHIVING: A VIEW FROM THE NATIONAL ARCHIVES OF AUSTRALIA

Adrian Cunningham

Paper for DigCurr2007 Conference, Chapel Hill, North Carolina

Harold Macmillan once described the life of a Foreign Secretary as being ‘forever poised between a cliché and an indiscretion’. While it is certainly not my intention in this paper to add to the already vertiginous mountain of digital curation clichés and truisms, I cannot help but feel that much of what I am about to say is a case of stating the bleeding obvious. It is perhaps indiscreet therefore to argue that a lot of what should be bleeding obvious appears to have been overlooked amongst the millions of words that have been written on this topic in recent years. The time has come for the debate to revisit some of the first principals of archival endeavour.

In this paper I will review the past ten years of digital curation endeavours at the National Archives of Australia, placing those endeavours in the broader Australasian context and identifying the major challenges that still require resolution. It will not surprise you to hear that one of these challenges is securing access to the various skills and capabilities that are required for digital curation, Australian style.

But before taking you on an Australian journey, let me outline three key messages, the ones that all too rarely rate a mention in a digital curation discourse that is too often myopic and terminologically confused.

Three key messages

Key message 1

Just as archiving (the management of archives and records) is but one form of curation, so too is digital archiving just one form of digital curation. Yet the two terms are so often used interchangeably as to appear to be synonymous. They should not be.

Digital curation of archival materials is not just about digital collection management. In fact, I would argue that the curation of digital records is sufficiently distinct as a curatorial activity as to warrant the use of a different term – digital archiving. In making this claim I realise that I am swimming against a strong terminological tide. But as an archivist I am prepared to draw a line in the sand and say that I have had enough of my professional language being misappropriated, abused and twisted by others – a trend which began years ago when Information Technology professionals began talking about the ‘archiving’ of back-up tapes and the like. Apparently ‘archiving’ is now just a technological sub-routine, not a rich and complex professional endeavour in its own right. Lest you feel I am being too precious and separatist about this, let me make it clear that I applaud the inclusive digital curation mission and I believe absolutely that digital archivists should work within a broad collaborative cross-domain environment to share ideas and solve problems. But broad cross-domain collaboration does not serve us well if it means we ignore the vitally

important differences between our various professional missions. This leads me on to my second key message.

Key message 2

Digital archives are different from digital libraries.

Just as archives are different from libraries and museums, so too should digital archives be different from digital libraries and museums. At this point we need to remind ourselves why archives are different from libraries. It is worth quoting ‘The Archivists’ Mission’ from the Australian Society of Archivists:

Archivists ensure that records which have value as authentic evidence of administrative, corporate, cultural and intellectual activity are made, kept and used. The work of archivists is vital for ensuring organisational efficiency and accountability and for supporting understandings of Australian life through the management and retention of its personal, corporate and social memory.

The nature of archival materials (records) is fundamentally different from the nature of library collections. Records provide evidence of decisions and activities. They derive their meaning and value from the myriad of contextual relationships surrounding their creation and use – relationships that have to be documented and understood. This is the core business of archivists. Archivists document recordkeeping activity in order that valuable records can be carried forward across time and domains of use in ways that ensure that their meaning and utility persists. Because records are created within systems that support and enable human activity (be they business systems or recordkeeping systems, however rudimentary in design), in order to understand records as evidence of human activity it is necessary to understand how their systems of creation and use operated. One way of understanding the work of archives, therefore, is to say that archives implement and manage systems for carrying recordkeeping systems forward across time and domains of use. The peculiar challenge of archiving is devising and implementing strategies for preserving the evidential meaning of records by capturing and preserving records in context. This is achieved through complex, dynamic, interlocking and finely engineered metadata regimes. Recordkeeping metadata is fundamentally different to and infinitely more complex than resource discovery metadata and preservation metadata. It is event-oriented metadata in an object-oriented world.

Even though they did not use words such as metadata to describe their documentation systems, our predecessors nevertheless worked all this out some generations ago. They implemented impressive regimes for carrying non-digital archives and records forward through time in our archival programs. They understood that archives are different from libraries, not because we like to be exclusionist, but because of the fundamentally different challenges posed by the nature of the material that is the locus of our work. Yet in the digital age we seem to have forgotten these fundamentals. Digital archives are at risk of being managed just like vanilla digital libraries, thus dumbing down the peculiar challenges and complexities of preserving *records*.

Preserving individual digital objects in bulk is, these days, relatively easy. We have made enough progress with digital preservation in recent years that we can probably all agree on that. In fact, preserving decontextualised digital objects is orders of magnitude easier than preserving the evidential and contextual meaning of digital

records created within complex systems and work practices. What is not easy, and what we have not yet fully come to grips with, is developing and implementing comprehensive regimes for capturing and managing records as evidence in context from before the point of creation for as long as those records are required by their creators and by society at large. For as long as we continue to regard digital libraries and digital archives as synonymous, we will continue to fail to address this challenge.

Key message 3

Digital archiving requires active archival intervention across the entire records continuum.

In other words, digital archiving is not just end-of-life-cycle collection management. This brings us inevitably to the OAIS Reference Model. Within its limitations the OAIS model is a good model for managing digital libraries. The problem is that its limitations are not recognised. Instead, OAIS has been uncritically adopted by all digital curators as accommodating everything we ever need to know about digital curation. As the American journalist Walter Lippman once said, “When we all think alike, we are not thinking”.

The problem with the OAIS model is that it assumes that submission information packages are out there, and that they simply have to be found, described and ingested into our digital repositories. Yet, our recent experiences with recordkeeping in modern organisations refute this assumption fairly comprehensively. What we know is that organisations, for all their gigabytes of data, have lost the ability to make and manage accurate, authentic and meaningful records of their activities. If you ask most organisations nowadays what digital records they have, how they are managed and how long they need to be kept for, you will almost certainly be met with incomprehension. Not only are they probably unable to answer the question, more often than not they won’t even understand it. They might be able to tell you how much data they have, but they won’t know how many records they have, what these records are, and how important or trivial they might be. The relentless technological juggernaut has ridden right over the top of basic information management techniques and strategies. The OAIS model makes no attempt to address what is probably the biggest single challenge facing digital archivists. If digital curation is to be successful it has to include intervention in the creation and management of digital information, not just take submission information packages as a given and go from there. In short, ignoring the front end of records creation is a recipe for having no submission information packages that are worth ingesting. We will have lovely digital repositories that will contain nothing with any real meaning or value. We will have failed in our mission to document the important things that happen in society and in public administration.

So, with these three key messages in mind, how has the National Archives of Australia faced up to the digital archiving challenge?

Improving government recordkeeping

In some ways the mid-1990s were no different to today. Then, as now, people were inclined to view the digital archiving challenge as being purely a matter of devising workable approaches to digital preservation. What was different, though, about that

time was that we felt completely overwhelmed by the digital preservation challenge. It was perhaps understandable in those days (but much less so now) for people to regard the challenge in purely technical digital preservation terms, because that was the 'in your face' issue. Archival programs worldwide invested every dollar they could spare on researching digital preservation. In the face of this widespread alarm, the National Archives of Australia, took the odd step of pretty much ignoring digital preservation, at least for a few years.

This wasn't just wilful perversity, there was method to our madness. First, we looked at our available resources and decided that researching or experimenting with digital preservation was likely to be a bottomless pit. Better to let other people explore solutions and conduct experiments, so that we could learn from their experiences. The more important consideration, however, was the realisation that we needed to become much more actively engaged in influencing recordmaking and recordkeeping in government agencies. Despite the absence of a strong legislative mandate or additional funding, we effectively took on a new function – that of being a recordkeeping standards setter and expert advisor. We considered that this was the more critical issue to address with our limited resources. Until we felt that agencies had regimes in place for making and keeping good digital records, there was no point in investing effort in developing a digital preservation program. It was a case of setting priorities and dealing with first things first.

This decision placed major strains on the organisation. Taking on a whole new function is never easy. Staff needed to embrace non-traditional concepts, strategies and modes of operation. Sitting within our comfort zone behind the walls of the repository doing business as usual was recognised as the fast route to oblivion. In 1994 we upset many of our professional and agency colleagues by announcing a distributed custody policy for electronic records. In effect, we were admitting that we were unable to manage electronic records in archival custody, so there was no point in agencies transferring such records to us – if indeed they had any to transfer. We decided that it was better for the records to stay in the custody of the agency that had the business need for and the technical expertise to manage the records in the first place. While this may have looked as though we were reneging on our archival responsibilities, it at least reflected an honest assessment of our capabilities at the time. The distributed custody policy gave us the time and the space to reinvent ourselves as recordkeeping standards setters and advisors.

The first fruits of this work came in 1996 when Standards Australia published the world's first national standard for records management, AS 4390. This standard, which was the fruit of a truly national collaborative effort, provided the basis for the later ISO standard, ISO 15489. As an aspirational 'best practice' standard, rather than a reflection of any current practices, AS 4390 gave us the high level blueprint for what we needed to implement across the entire Australian Government. But before we could attempt to change the entire Australian Government we had to change ourselves. Most of our staff were completely unfamiliar with records continuum thinking, the thinking that was embodied in AS 4390. In 1998 Monash University, the spiritual home of records continuum theory, was contracted to deliver a year-long training course in modern recordkeeping theory and practice to NAA staff. All staff above a certain classification were given two days a week on work time for 12 months to pursue the education delivered by Monash over the Internet. Many of the NAA's

standard operations and services were suspended in order to free up staff time for the Monash training. This suspension of normal business was itself an important circuit breaker in moving from the old regime to a new regime. By the end of that year staff were equipped with the conceptual knowledge and enthusiasm that was needed for the NAA to reinvent itself as a recordkeeping standards setter and expert advisor.

The initial fruits of the reinvention were unveiled in 2000 with the release on the NAA website of the *e-permanence* suite of modern recordkeeping standards and guidelines, the foundations of which were the DIRKS (Designing and Implementing Recordkeeping Systems) methodology/manual and our recordkeeping metadata standard. Included in the suite were guidelines on functional analysis and classification, guidelines for archiving web-based records and a variety of training materials. Since 2000 this suite of modern recordkeeping tools and guidelines has been continuously expanded, fine-tuned, revised and reshaped to reflect the changing recordkeeping realities of government and the lessons we have learnt during the implementation process. This is a never ending process, not just because the world never stands still, but also because of the ongoing challenge of turning theoretical models and frameworks into practical advice that can be adapted for the wide variety of circumstances faced by government agencies, large and small.

It is one thing to develop recordkeeping standards and guidelines. It is quite another to get government agencies to take notice of them, understand them and implement them. Recordkeeping is never going to be a sexy attention grabber in government, except perhaps when things go disastrously wrong and poor recordkeeping is identified (as it usually is) as a major contributing factor to failures in public administration. The NAA is small and has limited influence, while the Australian Government is large and complex. On their own, promotional and training strategies only get you so far down the path of whole of government change management. To really succeed archivists need strong allies, such as the head of the public service and the Auditor-General. Perhaps the biggest single factor in getting Australian Government agencies to take recordkeeping seriously has been the activism of the Auditor-General. Auditors are natural allies for archivists, because they absolutely understand the importance of good records. Since 2002 the Australian National Audit Office has conducted three separate audits of recordkeeping in Australian Government agencies, the results of which have been sobering to say the least. Heads of agencies pay close attention to published audit reports, much more so than they will ever pay to the messages coming out of the National Archives. The combination of agency readiness to transform their recordkeeping systems from paper to digital (something that took quite a bit longer than we originally estimated) and the heightened administrative attention being given to recordkeeping has now finally made recordkeeping one of the major topics of bureaucratic discussion in the Australian Public Service.

As of 2007 we still have a very long way to go to achieve recordkeeping nirvana in government agencies. In fact, I doubt if we ever will achieve this nirvana. In this day and age working with government to improve its recordkeeping is complex, frustrating and difficult. But we have no option but to keep trying. Improvements in one agency will probably be matched by deteriorations in other agencies. Nevertheless, we have to keep striving for continuous improvements, while developing strategies for coping with recordkeeping imperfection. We cannot expect

perfect recordkeeping, but we can expect agencies to take the issue seriously and to address the high risk/high significance areas of their operations with recordkeeping strategies that are achievable, sustainable and fit for purpose.

Digital preservation project

Following the launch of *e-permanence* in 2000, the NAA decided that it could set aside some resources to address the long postponed issue of digital preservation. As a result, the Agencies to Researcher digital preservation project was instituted in 2001. The first stage of the project was to research approaches to digital preservation from around the world and to devise an approach or mix of approaches that would be suitable for the NAA. This work culminated in 2002 with the release of a Green Paper ‘An Approach to the Preservation of Digital Records’.¹ The Green Paper argued that digital records are performances – the result of an interaction between data and technology. The preservation imperative, therefore, is not so much one of preserving the data, as of preserving the ability to recreate the performance in a way that accurately and authentically replicates the essential aspects of the user’s experience of the record.

In operationalising the approach presented in the Green Paper the NAA decided to avoid reliance on regular software migrations across proprietary platforms. Instead we opted for a strategy of ‘normalising’ records created in proprietary software applications and file formats into openly documented archival file formats and linking those objects to the necessary contextual and descriptive metadata. In the case of text-based records, the archival file formats and all of the metadata would be encoded in XML. A suite of open source software tools and plug-ins² called Xena (XML Electronic Normalisation of Archives) was developed for normalising and then re-rendering for use records originally created in proprietary formats. At the same time a suite of open source software tools for performing and documenting digital preservation activities was also developed. While these tools were developed primarily for use within the archival repository, the NAA has also developed “Xena-lite” for government agencies and other organisations that need to preserve the digital records that they need to retain in their own custody. All of these tools are available for inspection and download on SourceForge. Because it is open source, anyone anywhere in the world who has some Java programming skills can use, extend or enhance the Xena source code. Indeed, the NAA welcomes global community collaboration of the kind embodied in the open source movement.

The NAA now has a fully functioning, secure offline digital repository and is accepting and processing transfers of born-digital archival-value records from agencies. Nevertheless, we regard this work as still being at the cottage industry or proof of concept stage. We know that we need to be able to perform this work on an industrial scale for billions of records. We also know that we need to be able to provide greater support for digital preservation work in those agencies that need to preserve long term temporary value (ie. not archival value) born-digital records for a long time, in some cases for as long as 120 years. At present we simply do not have the capacity to perform all of this work at this scale, even though we are confident that

¹ http://www.naa.gov.au/recordkeeping/er/digital_preservation/summary.html

² For example some of the Xena plug-ins use the Open Document Format – ODF.

we know how to do it. We need our Government to recognise our needs in this area and to fund us over the transitional period during which we have to maintain dual archival operations for both paper and digital records.

Total end-to-end digital archiving

Having in place a regime for improving recordkeeping in creating agencies and a program for digital archives ingest and preservation are two important pieces of the digital archiving jigsaw. But we need more in order to be a fully functional digital archive. At first the NAA, I imagine like most archival programs, thought that doing digital preservation was all we needed to do to become a digital archive. It took two or three years, but the realisation eventually dawned upon us that the digital preservation project was not going to give us all the tools that we needed to perform end to end digital archiving. Just as there are more to archival operations than the preservation function, so there are more to digital archives than the digital preservation function.

In 2004 the NAA instituted a new project called MADIRA – Managing Digital Records for Access. MADIRA identified the remaining pieces of the digital archiving puzzle that the NAA needed to put into place before it could fulfil the original 2001 “Agencies to Researcher” vision. The Agencies to Archives bit of the process is in place – we can get digital records from agencies into our deep secure archival digital repository, and fully document all of our processes up to that point. Other archival functions, notably intellectual control/context description (once known as arrangement and description) and access management, still need to be put in place before we can deliver meaningful digital records as performances in context to our end-users. Because we have the luxury of a 30 year closed-access period for most of our holdings, this is perhaps not the most urgent priority that we are faced with. Nevertheless, we will have to address it before too long and – make no mistake – it will require serious resources and intellectual effort. In fact, at present, we are not at all sure where these resources are going to come from, but one way or another we will have to find them – perhaps in the form of additional funding from government.

Before leaving this quick overview of the NAA’s digital archiving endeavours I need to point out that the NAA is not the only player in the digital archiving space down under. In fact, the NAA is but one of ten different public records institutions in the different jurisdictions in Australia and New Zealand. As a small, some would say incestuous community, these ten institutions have a long history of collaborating and sharing ideas and best practices. In 2004 this habit of collaborating was given formal shape with the creation of the Australasian Digital Recordkeeping Initiative (ADRI). ADRI members are committed to working together to develop and implement a common Australasian approach to making, keeping and using digital records across the entire records continuum in each of the member jurisdictions. Limited resources are deployed collectively to work on priority joint projects and develop products that will be of benefit to all the ADRI member institutions. More information on ADRI can be found at <http://www.adri.gov.au>

Concluding thoughts on the skills and capabilities needed for digital archiving

Back in the early 1990s I had a colleague who regularly lamented just how much today's archivists need to know. At the time he was reacting to the influx of desktop computers, online networks and the increasing demands of public sector management reform – all of which added to pre-existing needs to know about archival theory and processes, Australian history, etc, etc. Well, things have not got any better since then – in fact they have probably got a lot more challenging.

Certainly, the NAA's experiment with the Monash University training course in 1998 was an early recognition that traditional skills and training did not provide adequate preparation for the challenges of digital archiving. While we can and must forge partnerships with other professions such as ICT, lawyers, business analysts, communications experts and educators, there are nevertheless a range of skills that every digital archivist needs today. So, by way of concluding this paper I will simply list some of the more important of those skills (in no particular order):

- Knowledge of the full range of recordkeeping theory and practice and the role of archives in society;
- Knowledge of how records sit in the broader information management landscape;
- Knowledge of the way modern organisations work, office processes, the machinery of government, etc;
- How to prepare business cases;
- Modelling and analytical ability (including functional and work process analysis);
- Communication, influencing and change management skills (get out of the basement and into the Board Room);
- Broad current affairs and historical knowledge;
- Systems design and implementation skills;
- ICT awareness and familiarity;
- Consultation and negotiation skills;
- Flexibility and good judgement;
- Knowledge of the workings of e-business and e-government;
- Research skills;
- Knowledge of metadata regimes for discovery, recordkeeping, data management, etc;
- Awareness of legal, regulatory and governance frameworks;
- Risk assessment and management skills;
- Knowledge of auditing and compliance assessment approaches and regimes;
- Knowledge of security management regimes in ICT, including encryption and authentication;
- Knowledge of broader digital curation communities and initiatives;
- XML awareness;
- Disaster preparedness, business continuity skills;
- Knowledge of approaches to quality control;
- Understanding of how to manage documentation of provenance and context in archival systems; and
- Knowledge of storage options and technologies.

Given the length of this list it is not at all surprising that there are very few (if any) individuals who can confidently claim to be a fully rounded digital archivist. Indeed, I sometimes wonder if our expectations can only be delivered by super-humans! Like the rest of the world, Australia is experiencing a chronic lack of digital archiving capabilities. While we have developed competency standards and capability frameworks, it is quite another thing to build and sustain the education and training infrastructure that is needed to develop these capabilities, especially for what is a boutique and not especially well remunerated occupation in a small country. Australia has just two accredited university level archival studies programs (Monash University in Melbourne and Edith Cowan University in Perth). Recently, the Australian National University in Canberra, in close consultation with the NAA, instituted a program called “A Systems Approach to the Management of Government Information”. Time will tell whether this promising development attracts sufficient enrolments to be a viable long-term source of the kinds of education that we so badly need.

Ultimately, digital archiving is not just an interesting source of research projects and academic theorising, there are pressing societal and organisational needs to develop sustainable industrial-scale digital archiving implementations in our national and state institutions. The NAA cannot yet claim to have such an implementation because of shortfalls in funding in available skills. Nevertheless, we are totally confident that we are pretty close to fulfilling this vision and that we have a detailed grasp on how such an implementation can work and what it will look like – notwithstanding the fact that digital archiving will forever remain a contested work in progress, rather than a settled orthodoxy.